Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 6007–6011, June 1994
Medical Sciences

Likelihood ratios for DNA identification

(population structure/DNA typing/forensic science)

A. COLLINS AND N. E. MORTON

CRC Genetic Epidemiology Research Group, Department of Child Health, University of Southampton, Level G, Princess Anne Hospital, Southampton, SO9 4HA, United Kingdom

Contributed by N. E. Morton, February 2, 1994

ABSTRACT **Likelihood ratio (LR) tests are provided for the three alternatives to DNA identity: exclusion, coincidence, and kinship. The coincidence test uses the radius of coalescence to conserve the observed frequency of single band phenotypes. Genotype probabilities under kinship are derived for mating groups, specified relatives, and structured populations; and unbiased estimates of the genetic parameters are provided. The LR is made robust to gene frequency errors by specifying the mean matching probability, and the tolerable loss of information this entails is determined by LR theory. This straightforward application of the seminal work of Jerzy Neyman and Sewall Wright strongly supports the use of LRs and kinship for presentation of DNA evidence by expert witnesses and committees.**

Neyman and Pearson (1) demonstrated that likelihood ratios (LRs) are the optimal basis for statistical decisions, whether or not there is an hypothesis about prior probabilities (1–3). Therefore LRs have Bayesian appeal but are not Bayesian. They have rich statistical properties and give a measure of divergence between hypotheses in information theory. Discussion of LRs by the National Research Council Committee on DNA Technology in Forensic Science (4) has fatal errors, leading to the "ceiling principle" that is neither a ceiling nor a principle (5–8).

Despite their advantages, LRs are not generally used for DNA identification. Molecular biologists developed an alternative procedure called match/binning whereby a match is declared by one criterion and its probability is calculated for a different one, using a cautious approach designed to favor the defendant (9). This has drawn criticism from statisticians because it can lead to ambiguity when two bands fall close together in different bins or far apart in the same bin, its properties are difficult to establish, and the superiority of LRs has been demonstrated. Adoption of LRs for DNA identification (with different but consistent LRs for exclusion, coincidence, and kinship) will increase efficiency and reliability, provide a rigorous solution to the search for conservative presentation of evidence, disarm criticism, and be more comprehensible to the court (10, 11). Advances in molecular techniques, especially recognition of discrete alleles, will alter error densities but not the fundamental logic.

Theory

DNA evidence E^j derived from locus j is the union of two pieces of evidence E^j_s and E^j_c contributed by individuals s and c called *suspect* (defendant) and *culprit* (an evidential or criminal sample). The population k of the culprit and the relationship R between them are relevant but usually unknown. Different hypotheses about k, R, the error density, and algorithms to estimate gene frequencies lead to different LRs. An expert witness can testify about these hypotheses and the support for each, but he should not usurp the responsibility of the court to determine the most credible hypothesis.

Let H_0 be a null hypothesis about the relationship of c and s and H_1 be an alternative hypothesis specifying a closer relationship. With k and R implicit, the general LR for the jth locus is

$$\lambda_j = \frac{P(E^j_c, E^j_s|H_1)}{P(E^j_c, E^j_s|H_0)} = \frac{P(E^j_c|E^j_s, H_1)}{P(E^j_c|E^j_s, H_0)}, \quad [1]$$

since the marginal probability $P(E^j_s)$ is independent of k and R. It is convenient to take the logarithm of odds (lod) $z_j = \ln \lambda_j$ so that $Z = \Sigma z_j$ is the evidence against H_0 in favor of H_1, the LR is $\lambda = e^Z$, and the probability of $\lambda \geq A$ under H_0 is less than $1/A$ for $A > 1$. Additivity (independence) of unlinked loci is justified theoretically because relationship does not induce linkage disequilibrium and may be confirmed empirically as zero correlation of lods. The divergence between H_1 and H_0 is defined as $E_1(Z) - E_0(Z)$, where $E_i(Z)$ is the mean lod when H_i is true (3). Three tests are defined against null hypotheses of exclusion, coincidence, and kinship.

In the *exclusion test*, H_0 denotes exclusion because of a different genotype, and H_1 is inclusion because of the same genotype. It is the LR analog of the match step in match/binning and is designed to protect the suspect against false inclusion, leaving to other tests the distinction among coincidence, kinship, and identity. It is adapted to the current standard of representing alleles by fragment length instead of sequence and would be unnecessary for an error-free determination of sequence or fragment length. Since the distributions under both hypotheses are continuous, there is no need to define bins or estimate genotype frequencies. Consider the jth locus at which the suspect has two fragments, the natural logarithms of which are $x^j_1 \leq x^j_2$. The corresponding logarithmic lengths in the culprit are $y^j_1 \leq y^j_2$. We specify logarithms to stabilize the variance when the standard deviation of replicates is roughly proportional to fragment size. Then the error density depends only on the distribution of samples from the same individual, $f(u_j, v_j)$, where $u_j = y^j_1 - x^j_1$ and $v_j = y^j_2 - x^j_2$ and orthogonal functions are $S_j = u_j + v_j$, $D_j = u_j - v_j$ (12). Since the vectors symbolized by x and y may be interchanged, $E(u) = E(v) = E(S) = E(D) = 0$ and $E(SD) = E(u^2) - E(v^2) \approx 0$, but $E(uv) > 0$ if there is uncorrected band shifting and

$$\sigma^2_S = E(S^2) = E(u^2) + E(v^2) + 2E(uv)$$

$$\sigma^2_D = E(D^2) = E(u^2) + E(v^2) - 2E(uv).$$

Being functions of four variables, each symmetrical and roughly normal, the distributions of S and D are nearly

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: lod, logarithm of odds; LR, likelihood ratio.

normal. Therefore the error density for replicates of the same genotype is

$$P(E_c^j|E_s^j, H_1) = \frac{1}{2\pi\sigma_{S_{j1}}\sigma_{D_{j1}}} e^{-(S_j^2/2\sigma_{S_{j1}}^2 + D_j^2/2\sigma_{D_{j1}}^2)}\, dy_1^j\, dy_2^j. \quad [2]$$

Under H_0 the distribution has approximately the same form with greater variances $\sigma_{S_{j0}}^2$ and $\sigma_{D_{j0}}^2$, although deviation from normality will be more pronounced if fragment size is multimodal (13). The differential elements cancel and so

$$z_j = \ln(\sigma_{S_{j0}}\sigma_{D_{j0}}/\sigma_{S_{j1}}\sigma_{D_{j1}}) + \tfrac{1}{2}[S_j^2(1/\sigma_{S_{j0}}^2 - 1/\sigma_{S_{j1}}^2) + D_j^2\,(1/\sigma_{D_{j0}}^2 - 1/\sigma_{D_{j1}}^2)]. \quad [3]$$

Substructure or other relationship between s and c is not considered in the exclusion test, since exclusion makes relationship irrelevant while inclusion leads to tests that distinguish among identity, coincidence, and kinship. The forensic population that gives the smallest lod provides the most conservative admissible test, but this source of variation is minor. If the LR is large (say >1000), the suspect cannot be excluded and coincidence should be tested. (At this point a Bayesian statistician would introduce prior probabilities and costs of different wrong decisions, but this is optional and controversial).

In the *coincidence test* H_0 denotes coincidence (a different random individual with the same phenotype) and therefore specifies a discrete variable that is conveniently represented by bins, each of which would contain only one sequence or fragment length in an error-free assay. Random mating (panmixia) is assumed. H_1 denotes the same individual, with the error density integrated over the bin. There are many ways to perform integration. Berry *et al.* (13) applied normal kernel smoothing, but choice of the smoothing parameter is arbitrary. Devlin *et al.* (14) assumed a model for size of flanking regions and number of repeats, which is necessarily approximate and not testable on current evidence. Spline or other interpolation might be used. Morton *et al.* (15) proposed $dy_1^j = dy_2^j = 2\hat{w}_j$, where $\hat{w}_j$ is the radius of coalescence for the jth locus, defined so that bins of size $2\hat{w}_j$ give the same estimate of homozygosity h_j as the observed frequency h_{oj} of individuals with single bands. They estimated $\hat{w}_j$ by fitting $h_j = 1 - e^{-cj^w}$, where c is a constant specific for race and locus, and taking $\hat{w}_j = -\ln(1 - h_{oj})/c_j$. The resultant bins are treated as alleles. Thus if y falls in the rth bin with frequency q_r, the probability of genotype G_rG_s (with locus j implicit) is

$$P(E_c|E_s, H_0) = \begin{cases} q_r^2 \text{ if } r = s \\ 2q_rq_s \text{ if } r \neq s. \end{cases} \quad [4]$$

In our experience $\hat{w}$ is approximately equal to three replicate standard deviations, and therefore the integral of Eq. 2 between $y \pm \hat{w}$ is nearly 1. Under these conditions, the coincidence LR is the reciprocal of the panmictic matching probability. The coincidence test is designed to protect the suspect against a chance match but does not protect against a related culprit. It allows greater choice than the exclusion test, since k may be varied, integration may be performed in different ways, and these genotype-specific matching probabilities may be replaced by their mean value (16, 17):

$$\bar{P}(E_c|E_s, H_0) = 2\left(\sum q_r^2\right)^2 - \sum q_r^4. \quad [5]$$

This loses some information [measured as $E_0(Z)$] but makes the calculation extremely robust with respect to sampling errors and choice of population. Unbiased estimates of the moments are given by

$$\sum q_r^2 = \frac{E\{\sum n_r(n_r - 1)\}}{N(N-1)} = \frac{\sum n_r^2 - N}{N(N-1)}$$

$$\sum q_r^3 = \frac{E\{\sum n_r(n_r - 1)(n_r - 2)\}}{N(N-1)(N-2)} = \frac{\sum n_r^3 - 3\sum n_r^2 + 2N}{N(N-1)(N-2)} \quad [6]$$

$$\sum q_r^4 = \frac{E\{\sum n_r(n_r - 1)(n_r - 2)(n_r - 3)\}}{N(N-1)(N-2)(N-3)} = \frac{\sum n_r^4 - 6\sum n_r^3 + 11\sum n_r^2 - 6N}{N(N-1)(N-2)(N-3)},$$

where n_r is the observed number in the rth bin of the chosen population and $N = \Sigma n_r$ (18). Although small samples do not give reliable estimates of bin frequencies, Eq. 6 avoids the error of attributing bias to substructure (19). If the LR is large (say >1000), coincidence is rejected and kinship should be tested.

The *kinship test* answers the most persistent and troublesome issue raised by the defense: might a match be due to close or remote relationship between suspect and culprit, where remote relationship may be called substructure? A valid answer makes it unnecessary to partition a large forensic sample into small subsamples or to investigate foreign populations. H_1 denotes the same individual as in the coincidence test and H_0 denotes a particular relationship. A large lod favors identity of suspect and culprit, even against the alternative of a related culprit. Naturally the information (measured as divergence) is less than that for exclusion and coincidence tests, and strong evidence may require testing of additional loci or specific relatives. However, if no relative is under suspicion, a moderately large LR (say >100) provides strong evidence (not proof) of identity.

Several kinship models may be entertained (20). In the progenitive (parent–child) model, one allele is identical by descent and so

$$P(E_c|E_s, H_0) = \begin{cases} q_r \text{ if } r = s \\ (q_r + q_s)/2 \text{ if } r \neq s. \end{cases} \quad [7]$$

As in Eq. 5 we may replace genotype-specific matching probabilities by their mean value (16, 17):

$$\bar{P}(E_c|E_s, H_0) = \sum q_r^2. \quad [8]$$

These results are the special case $\varphi = \frac{1}{4}$ of regular (noninbred) unilineal relatives with kinship φ for which

$$P(E_c|E_s, H_0) = \begin{cases} 4\varphi q_r + (1 - 4\varphi)q_r^2 \text{ if } r = s \\ 2\varphi(q_r + q_s) + (1 - 4\varphi)(2q_rq_s) \text{ if } r \neq s \end{cases} \quad [9]$$

(21). The mean matching probability is

$$\bar{P}(E_c|E_s, H_0) = 4\varphi\sum q_r^2 + (1 - 4\varphi)\left[2\left(\sum q_r^2\right)^2 - \sum q_r^4\right] \approx 2\sum q_r^2\left(\sum q_r^2 + 2\varphi\right). \quad [10]$$

In the affinal model the culprit is related as closely to the suspect as a spouse would be (20, 22). This allows for possible inbreeding of either individual. Then up to terms in α^2,

$$P(E_c|E_s, H_0) = \begin{cases} q_r^2 + 5q_r(1 - q_r)\alpha \text{ if } r = s \\ 2q_rq_s\left[1 + \dfrac{q_r + q_s - 5q_rq_s}{q_rq_s}\,\alpha\right] \text{ if } r \neq s, \end{cases} \quad [11]$$

where α is the mean inbreeding coefficient in the population to which both belong. There is considerable evidence about

values of α in different populations (20). Higher order terms are negligible and depend on the unknown distribution of gene frequencies among mating groups. The mean matching probability is

$$\bar{P}(E_c|E_s, H_0) = 2\left(\sum q_r^2\right)^2 - \sum q_r^4 + 2\alpha\left[2\sum q_r^2 + \sum q_r^3 + 3\sum q_r^4 - 6\left(\sum q_r^2\right)^2\right]$$

$$\approx 2\sum q_r^2\left(\sum q_r^2 + 2\alpha\right). \quad [12]$$

Sibs are the most common bilineal relatives. If they are not inbred the matching probability (21, 23) is

$$P(E_c|E_s, H_0) = \begin{cases}(1 + 2q_r + q_r^2)/4 \text{ if } r = s\\(1 + q_r + q_s + 2q_rq_s)/4 \text{ if } r \neq s,\end{cases} \quad [13]$$

with mean (16)

$$\bar{P}(E_c|E_s, H_0) = \left[1 + 2\sum q_r^2 + 2\left(\sum q_r^2\right)^2 - \sum q_r^4\right]/4. \quad [14]$$

This is a special case of regular (noninbred) bilineal relatives with probability c_p of having p genes identical by descent and conditional matching probability t_p, where

$$t_0 = \begin{cases}q_r^2 \text{ if } r = s\\2q_rq_s \text{ if } r \neq s\end{cases}$$

$$t_1 = \begin{cases}q_r \text{ if } r = s\\(q_r + q_s)/2 \text{ if } r \neq s\end{cases}$$

$$t_2 = 1$$

$$\bar{t}_0 = 2\left(\sum q_r^2\right)^2 - \sum q_r^4$$

$$\bar{t}_1 = \sum q_r^2 \quad [15]$$

$$\bar{t}_2 = 1$$

$$P(E_c|E_s, H_0) = \sum c_p t_p$$

$$\bar{P}(E_c|E_s, H_0) = \sum c_p \bar{t}_p$$

$$\varphi = c_2/2 + c_1/4.$$

There is considerable variation among LRs for the kinship test, depending largely on the model and magnitude of kinship (Table 1). The appropriate choice rests with the court, not the expert witness, who should, however, be well versed in the evidence on human population structure and its implications for the kinship test.

Examples

A trial of alternative algorithms requires a sample of replicate tests, a forensic data base of different individuals, and pairs representing suspect and culprit. The first two are provided by the Federal Bureau of Investigation for five loci typed with *Hae* III (9) and by Lifecodes Corporation for four loci typed with *Pst* I (24). We are grateful to Bruce Budowle and Ivan Balazs for these data bases and helpful criticism. For each forensic sample, replicates were generated by the same protocol on different gels, usually on different days and in different laboratories as part of a blind quality control. This material was analyzed by the 4N6 program, which performs calculations for LRs, kinship, and ancillary tests (15). Suspect and culprit pairs were generated for each locus by cyclical permutation within a specified population (and within individuals for replicates) after shuffling into random order. Thus N informative observations generate N pairs in which each observation appears once as a suspect and once as a culprit.

In Table 2 the locus with the smallest amount of information in the coincidence test is *D17S79*, which may have the highest frequency of null alleles by *Hae* III (25), leading to a high estimate of the radius of coalescence and therefore large bins. This loses information but does not simulate kinship. For a given probe, slightly more information is extracted by the 4-base cutter *Hae* III than by the 6-base cutter *Pst* I, which produces larger fragments. Since relative errors are to a first approximation uniform, large fragments tend to have large absolute errors, reflected by greater bin size measured in base pairs and correspondingly higher random matching probabilities.

Information to exclude a suspect is greater than for the coincidence test, since two bands far apart in the same bin favor exclusion. Despite error in fragment lengths, an innocent suspect is assured of exclusion if enough tests are performed. Formally stated, the probability that a random suspect is not excluded is less than $1 - \Pi_j(1 - M^j)$, where M^j is the probability of a coincidental match at the jth locus (20). The empirical matching probability is the frequency of exclusion LRs >1 in cyclical pairs. Even by a moderate battery of tests on unlinked loci, the probability that a random suspect is not excluded is less than 10^{-18} (Table 2).

The efficiency of the coincidence test as the ratio of expected lods is 0.89 when the mean matching probability is used. The efficiency rises to 0.95 for the kinship test. Many courts will consider the loss of information a reasonable price to pay for protecting the suspect against sampling errors and choice of an inappropriate population.

The kinship test offers even greater protection against substructure or other relationship between suspect and culprit. We have illustrated this by the unilineal model with $\varphi =$

Table 1. Coefficients of identity for regular relatives

Relationship	Kinship (ϕ)	Identity coefficients c_2	c_1	c_0	Degree of kinship (k)
Identical twins	1/2	1	0	0	0
Sibs	1/4	1/4	1/2	1/4	1
Double first cousins	1/8	1/16	6/16	9/16	2
Parent–child	1/4	0	1	0	1
Grandparent–grandchild (= uncle–niece = half sibs)	1/8	0	1/2	1/2	2
First cousins (= great grandparent–great grandchild = great uncle–niece)	1/16	0	1/4	3/4	3
First cousins once removed	1/32	0	1/8	7/8	4
Second cousins	1/64	0	1/16	15/16	5
Equal bilineal	$(1/2)^{k+1}$	$4\varphi^2$	$4\varphi(1 - 2\varphi)$	$(1 - 2\varphi)^2$	k
Unilineal	$(1/2)^{k+1}$	0	4φ	$1 - 4\varphi$	k

Table 2. Expected lods for LR tests (common logarithms, pooled ethnic groups)

Locus	Ref.	Exclusion		Coincidence, $E_0(Z)$		Kinship, $E_0(Z)$, $\varphi = 0.05$	
		$E_1(Z)$	$E_0(Z)$	Eq. 4	Eq. 5	Eq. 9	Eq. 10
D2S44	9	3.18	−1836	2.30	2.13	1.79	1.73
D1S7	9	2.95	−936	2.49	2.35	1.92	1.88
D17S79	9	2.55	−420	1.42	1.23	1.16	1.08
D4S139	9	2.61	−473	2.23	2.05	1.74	1.68
D10S28	9	3.36*	−1079*	2.17	2.06	1.72	1.69
D2S44	24	2.58	−467	1.90	1.65	1.50	1.40
D17S79	24	2.36	−274	1.74	1.46	1.37	1.26
D14S13	24	—*	−1589*	2.29	1.88	1.71	1.55
D18S27	24	2.49	−358	1.54	1.30	1.24	1.13
Total		—	—	18.08	16.11	14.15	13.40

*Few replicates for *D10S28* and no replicates for *D14S13*. Means of standard deviations from same reference were substituted. This affects only the exclusion test.

0.05, an absurdly high value for most forensic populations but appropriate under high levels of inbreeding. The efficiency declines to 78% for specific matching probabilities and to 83% for mean matching probabilities, but the decline is much less for more typical values of φ (15). The lost information can of course be recovered by testing more loci, and the court must decide on a reasonable balance between cost and credibility.

Discussion

Inevitably statistical methods for DNA identification lag behind advances in molecular biology, especially in the initial phases. Data bases are relatively small and not well designed, and there has been little quality control of calculations. Given the superiority of LRs over other statistics, what are the objections and complications?

The *exclusion test* depends on the densities of fragment lengths in replicates (error) and the general population. Some information is lost when the allelic distribution is neglected by considering only differences in fragment size, but the number of loci tested should be large enough to make this negligible. Error is minimized if the fragment lengths are well separated as in short tandem repeats or if a standard is run in every lane with differential labeling. Manual protocols use fewer standards for each gel, permitting band shifting and adding an extra source of variation to small fragments. Small PCR fragments are less subject to degradation than large restriction fragment length polymorphism fragments. Measurement error can be eliminated by procedures that resolve 1-bp differences, and then the logarithmic transformation of fragment lengths becomes unnecessary and the exclusion test becomes categorical.

The *coincidence test* will always be subject to dispute about the relevant population and sampling errors. Dispute is minimized by using mean matching probabilities and a large sample of the major ethnic group to which the suspect belongs: as Eq. 1 shows, this is a courtesy to the defendant and not a logical inference. Clearly the LR to reject coincidence must be set so high (by testing a sufficient number of loci) that the choice of sample is not critical. Usually three loci provide adequate evidence, although a few more are desirable. Any question about the propriety of multiplying locus-matching probabilities should be referred to the kinship test.

The distinction between coincidence and kinship is blurred on the hypothesis of random sampling from a subdivided population so that suspect and culprit are both inbred to extent α, but there is no kinship between them. This has been approached through the $2p$ rule that falsifies the frequency of homozygotes without modifying heterozygote frequencies (26), a substitute for kinship that makes the calculations no longer probabilities, invalidates LRs, and at usual levels of inbreeding is opposite to the biological effect it attempts to model. The mean matching probability is

$$\begin{aligned}\overline{P}(E_c|E_s, H_0) &= \sum_r [q_r^2 + q_r(1-q_r)\alpha]^2 + \sum_{r<s} [2q_rq_s(1-\alpha)]^2 \\ &= 2\left(\sum q_r^2\right)^2 - \sum q_r^4 + 2\alpha\left[\sum q_r^3 + \sum q_r^4\right. \qquad [16] \\ &\quad \left. - 2\left(\sum q_r^2\right)^2\right] + \alpha^2\left[\sum q_r^2 - 2\sum q_r^3 - \sum q_r^4 + 2\left(\sum q_r^2\right)^2\right].\end{aligned}$$

Hypervariable loci come close to the ideal system, which has $\Sigma q_r^m = 1/n^{m-1}$, where $m = 1, 2, \ldots$ and $n = 1/\Sigma q_r^2$ is the effective number of alleles. Assuming this, $\overline{P} \approx (2 - 2\alpha + n\alpha^2)/n^2$, which declines as α goes from 0 to $1/n$ and increases monotonically thereafter. Not until α reaches $2/n$ is the matching probability as great as under random mating. Since values of α in excess of $2/n$ occur only with strong preferential consanguineous mating, the assumption of $\alpha = 0$ in the coincidence test exaggerates the matching probability and is therefore favorable to the suspect without fudging homozygote frequencies by the $2p$ rule, which is inappropriate since kinship between suspect and culprit increases matching probabilities for heterozygotes as well as homozygotes. This completes the argument that kinship should not be incorporated into the coincidence test and that the $2p$ rule should not be used in any LR test.

There is an interesting distinction between the effect of B, the number of individuals that *must* include the culprit, and b, the number of individuals of known phenotype that *may* include the culprit. A crime aboard ship or on a desert island invokes B, which does not enter into the LR (27). Trawling a forensic data base for matches on the hypothesis of recidivism invokes b, the number of individuals trawled. Suppose m of these fail the exclusion test on the basis of information in the data base, and further investigation (including failure of exclusion on additional loci) identifies one of these as the suspect. Let C_d be the matching probability based on the data base and C_s be the matching probability based on loci tested subsequently. Then an appropriate joint matching probability in the coincidence test is $[1 - (1 - C_d)^b]C_s$. This Bonferroni correction is almost b times as favorable to the suspect as the uncorrected matching probability C_dC_s, but again the number of loci tested should make the difference negligible.

Jeffreys *et al.* (28) have suggested an extension of multilocus tests that uses oligonucleotides within a variable number of tandem repeats locus. This avoids some serious technical problems with multilocus restriction fragment length polymorphisms, but unless haplotypes are resolved, the weight of evidence (using empirical mean matching probabilities) is less than for tests based on alleles or haplotypes at multiple, unlinked loci.

The *kinship test* inherits all controversy about population structure. Since kinship does not cause linkage disequilibrium, matching probabilities are multiplicative over loci, providing the correct gene frequencies, kinship, and sampling model are used. The more emphasis there is on kinship, the less reason there is to question the multiplicative rule over loci, although methods are available to incorporate dependence whether caused by sampling of relatives or replicates or not (ref. 20; Eq. 15). With rare exceptions, it is plausible to assume that kin of an individual belong to his ethnic group, and so the choice of gene frequencies is limited by the forensic samples to which the suspect or culprit might reasonably be assigned (including the total data base). After long periods of neglect, the affinal model has come into favor to represent an isolate (local population or unusual ethnic

group) for which there is no large sample. Often there will be no reason to assume that suspect and culprit come from the same local population unless the suspect is guilty. However, kinship is an appealing hypothesis for the defense, and an expert witness should make appropriate calculations. Yasuda (22) derived probabilities for mates assuming that terms in α^2 are negligible, where α is the mean inbreeding coefficient, and Morton (ref. 20; Eq. 11) followed him in deriving conditional probabilities. This is justified even if α is as great as 0.05, which is exceptional. Moreover, the signs of coefficients of α^2 and α^3 depend on the unknown distribution of gene frequencies among (abstract) mating groups, which in turn depends on the vector of evolutionary sizes, the matrix of migration rates, and historical factors. Real populations are mixtures of local populations with different sizes and kinship. A beta distribution is plausible only for identical replicates under an island model (29), and its use may well be less accurate than the linear approximation although the effect in practice is trivial (30).

The most general criticism of kinship models is that they rest on the assumption of random drift among local populations within a region, but this objection is not likely to be raised either by advocates of neutral mutation or by others who recognize that migration dominates selection over regions that are large compared with neighborhood size (29). Genotype probabilities are functions of gene frequencies and kinship only in expectation, which does not apply exactly to any particular subpopulation. The same objections could be made to the Hardy–Weinberg law as a special case: gametes do not unite at random but through mating pairs with variable numbers of offspring, which give expected genotype probabilities only in the limit for large populations. The appeal of kinship theory is that no more accurate representation of genotype frequencies has been found. Given a sufficient number of tested loci, it is unlikely that different calculations based on sound theory and evidence would ever lead to different inferences, either in an ethnic group or a local population.

Forensic use of DNA has passed through three stages. The first used multilocus probes, which have high power, but present technical and statistical problems that restrict their use (31). The second stage used single locus probes with eclectic matching rules (32). The third phase applied the ceiling principle, which has proven indefensible, since possible relationship between suspect and culprit introduces kinship coefficients that cannot be modeled by falsifying gene frequencies (4). We are now in the fourth stage, in which the PCR is adopted, errors in fragments lengths can be eliminated, nonexpressed loci are chosen to show little variation among populations, matching probabilities are replaced by the richer armamentarium of LRs, the relevant hypotheses are recognized to be multiple (exclusion, coincidence, and kinship), and possible relationship between suspect and culprit is appropriately modeled.

A computer program for forensic use of DNA must address several aspects not considered here but incorporated in the 4N6 program. They include estimating kinship among forensic populations, testing for Hardy–Weinberg proportions, making pairwise disequilibrium tests, simulating relationship, matching a suspect against the data base, and making a case report on a suspect–culprit pair. The relevant theory has been published (20), and further applications will be made. However, progress in this field would be accelerated if its practitioners adopted the approach of laboratory workers, who welcome, instead of resisting, technical advances and quality control. Since population genetic theory is not controversial, future effort should be directed to analysis of forensic populations and presentation of results in ways that are at once most powerful, most reliable, and most readily comprehended.

LR theory was established in 1928 and essentially complete by 1959 (1–3). Kinship theory was introduced in 1921 (33) and matured by 1968 (22). The pioneers, Jerzy Neyman and Sewall Wright, were members of the National Academy, and the importance of their work is generally recognized. Nevertheless, these advances made when America led the world in statistics and population genetics are neglected by American courts and the National Research Council Committee on DNA Technology in Forensic Science, either through ignorance or in the belief that it is better to disregard the darkness than to light a candle. British courts accept LRs and kinship, which the London Metropolitan Police use routinely (albeit imperfectly) for presentation of evidence (34). American law must also be reconciled to science.

Note Added in Proof. Eq. 16 has recently been derived independently and discussed in detail (35).

1. Neyman, J. & Pearson, E. S. (1928) *Biometrika* **20**, 175–240; 263–294.
2. Wald, A. (1947) *Sequential Analysis* (Wiley, New York).
3. Kullback, S. (1959) *Information Theory and Statistics* (Wiley, New York).
4. National Research Council (1992) *DNA Technology in Forensic Science* (Natl. Acad. Sci., Washington, DC).
5. Cohen, J. (1992) *Am. J. Hum. Genet.* **51**, 1165–1167.
6. Weir, B. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 11654–11659.
7. Morton, N. E. (1993) *Eur. J. Hum. Genet.* **1**, 172–178.
8. Devlin, B., Risch, N. & Roeder, K. (1993) *Science* **259**, 748–749, 837.
9. Budowle, B., Giusti, A. M., Waye, J. S., Baechtel, F. S., Fourney, R. M., Adams, D. E., Presley, L. A., Deadman, H. A. & Monson, K. L. (1991) *Am. J. Hum. Genet.* **48**, 841–855.
10. Chernoff, H. (1992) *Stat. Sci.* **6**, 192–196.
11. Geisser, S. (1992) *J. Am. Stat. Assoc.* **87**, 607–614.
12. Buckleton, J., Walsh, K. A. J. & Triggs, C. M. (1991) *J. Forensic Sci. Soc.* **31**, 353–363.
13. Berry, D. A., Evett, I. N. & Pinchin, R. (1992) *Appl. Stat.* **41**, 499–531.
14. Devlin, B., Risch, N. & Roeder, K. (1991) *Am. J. Hum. Genet.* **48**, 662–676.
15. Morton, N. E., Collins, A. & Balazs, I. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1892–1896.
16. Lange, K. (1987) *Am. J. Hum. Genet.* **39**, 148–150.
17. Bishop, D. T. & Williamson, J. A. (1990) *Am. J. Hum. Genet.* **46**, 254–265.
18. Morton, N. E., Yee, S., Harris, D. E. & Lew, R. (1971) *Theor. Popul. Biol.* **2**, 507–524.
19. Krane, D. E., Allen, D. L., Sawyer, S. A., Petrov, D. A. & Hartl, D. L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10583–10587.
20. Morton, N. E. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2556–2560.
21. Cotterman, C. W. (1940) *A Calculus for Statistico-Genetics* (Ohio State Univ., Columbus).
22. Yasuda, N. (1968) *Am. J. Hum. Genet.* **20**, 1–23.
23. Evett, I. W. (1992) *J. Forensic Sci. Soc.* **32**, 5–14.
24. Balazs, I., Baird, M., Clyne, M. & Meade, E. (1989) *Am. J. Hum. Genet.* **44**, 182–190.
25. Chakraborty, R., de Andrade, M., Daiger, S. P. & Budowle, B. (1992) *Ann. Hum. Genet.* **56**, 45–57.
26. Chakraborty, R. & Zhong, Y. (1994) *Hum. Hered.* **44**, 1–9.
27. Balding, D. J. & Donnelly, P. (1994) *J. R. Stat. Soc. A* **157**, in press.
28. Jeffreys, A. J., MacLeod, A., Tamaki, K., Neil, N. J. & Monckton, D. G. (1991) *Nature (London)* **354**, 204–209.
29. Wright, S. (1951) *Ann. Eugen.* **15**, 323–354.
30. Balding, D. J. & Nichols, R. A. (1993) *Forensic Sci. Int.* **64**, 125–140.
31. Jeffreys, A. J., Turner, M. & Debenham, P. (1991) *Am. J. Hum. Genet.* **48**, 824–840.
32. Lander, E. S. (1991) *Am. J. Hum. Genet.* **48**, 819–823.
33. Wright, S. (1921) *Genetics* **6**, 111–178.
34. Evett, I. W. (1992) *J. Forensic Sci. Soc.* **32**, 5–14.
35. Li, C. C. & Chakravarti, A. (1994) *Hum. Hered.* **44**, 100–109.